УДК 519.652
DOI: 10.33764/2618-981X-2019-4-1-64-67

# О ПАРАМЕТРИЗАЦИИ ВЫБОРА ЧИСЛА ЗНАЧАЩИХ КЛАСТЕРОВ

***Павел Алексеевич Ким***
Институт вычислительной математики и математической геофизики СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Лаврентьева, 6, кандидат физико-математических наук, доцент, старший научный сотрудник лаборатории обработки изображений, тел. (383)330-73-32, e-mail: kim@ooi.sscc.ru

Одной из фундаментальных задач кластерного анализа является разбиение выборок многомерных данных на группы кластеров – объектов, близких между собой, в смысле некоторой заданной меры сходства. В ряде задач число кластеров задается исходно из физического смысла задачи, но чаще требуется определить их количество в ходе решения кластеризации. При большом числе кластеров, особенно, если данные «зашумлены», смысл задачи становится трудно воспринимаемым для анализа экспертами, и тогда искусственно уменьшают требуемое число кластеров рассмотрения. В работе представлены формальные средства слияния «соседних» кластеров в объединенный кластер, создающие основу для параметризации числа значащих кластеров в модели «естественной» кластеризации [1].

**Ключевые слова:** естественная кластеризация, слияние кластеров.

# ABOUT PARAMETRIZATION OF SELECTION OF SIGNIFICANT CLUSTERS

***Pavel A. Kim***
Institute of the Computational Mathematics and Mathematical Geophysics SB RAS, 6, Prospect Akademik Lavrentiev St., Novosibirsk, 630090, Russia, Ph. D., Senior Researcher of Image Processing Laboratory, phone: (383)330-73-32, e-mail: kim@ooi.sscc.ru

One of the fundamental tasks of cluster analysis is the partitioning of multidimensional data samples into groups of clusters – objects, which are closed in the sense of some given measure of similarity. In a some of problems, the number of clusters is set a priori, but more often it is required to determine them in the course of solving clustering. With a large number of clusters, especially if the data is "noisy," the task becomes difficult for analyzing by experts, so it is artificially reduces the number of consideration clusters. The formal means of merging the "neighboring" clusters are considered, creating the basis for parameterizing the number of significant clusters in the "natural" clustering model [1].

**Key words:** natural clustering, cluster merging.

## *Введение*

Проведение кластеризации – то есть разделение пространства объектов на классы, относится к важным практическим действиям, в процессе осуществления которых не только решаются массовые задачи, но и создаются автоматизированные обучаемые алгоритмы, с привлечением экспертов-учителей или же без них. Интуитивная постановка вопроса выделения границ разделяемых классов, как «близость» объектов внутри кластера и «удаленность» к объектам других кластеров, может иметь разные интерпретации, что приводит к существенному различию результатов, порождаемому такими разнородными алгоритма-

ми, используемых для решения одних и тех же задач. Важную роль в экспертизе решений играют специалисты-предметники, обеспечивающие целенаправленность продвижения к развитию прикладных библиотек кластеризующих алгоритмов [2]. При большом числе кластеров, особенно, если данные «зашумлены», смысл задачи становится трудно воспринимаемым для анализа экспертами, и тогда искусственно уменьшают требуемое число кластеров рассмотрения. В работе представлены формальные средства слияния «соседних» кластеров в объединенный кластер, создающие основу для параметризации числа значащих кластеров в модели «естественной» кластеризации.

## *Методы и материалы*

Теоретическую основу кластерного подхода к классификации дают диаграммы Вороного [3], которые могут работать не только для плоскости, но и в пространстве, разбивая пространство объектов на близкие между собой выбранные подмножества объектов.

Для естественной кластеризации опорными точками кластеров являются точки локальных минимумов высот поверхности, при этом границами кластеров становятся урезы перетоков водостоков [4]. Для поверхности рельефа дна мирового океана (рис. 1) мы можем выделить огромное множество точек локальных минимумов, однако для практических целей используются известные географические водные бассейны, существенно ограничивая число значащих кластеров. Это уменьшение обеспечивается выбором нулевого уровня водной поверхности мирового океана, сливая воедино, затапливаемые соседние кластеры в объемлющий кластер. Выделяемые географически моря в смысле нашей модели кластеры не образуют, зато кластерами выделяются озера, например, Байкал или Каспий. Отметим, что поскольку исходные кластеры образуются замкнутыми контурами Жордана, делящими поверхность на внутреннюю и внешнюю части, то при слиянии структура вложенных кластеров сохранит аналогичные разделимые части.

Характер слияния кластеров следует естественным механизмам слияния водостоков. По гидрографическому признаку различают три типа водохранилищ: *русловые*, *озерные и смешанные*. Водохранилище, которое образуется в результате преграждения течения реки плотиной и затопления речной долины, называется *русловым* (рис. 2, *а*). Такие водохранилища обычно имеют большую длину и площадь водного зеркала. Для создания в них больших запасов воды необходимо значительное повышение уровня воды.

*Озерное* водохранилище образуется в результате преграждения плотиной истока реки, вытекающей из озера (рис. 2, *б*). Вода при этом заполняет озерную чашу. В таких водохранилищах с большой площадью водного зеркала могут создаваться значительные запасы воды при сравнительно небольших повышениях уровня озера. При возведении плотины несколько ниже истока реки, вытекающей из озера, образуется *смешанное* водохранилище, которое включает емкости чаши озера и прилегающей к нему долины реки (рис. 2, *в*).

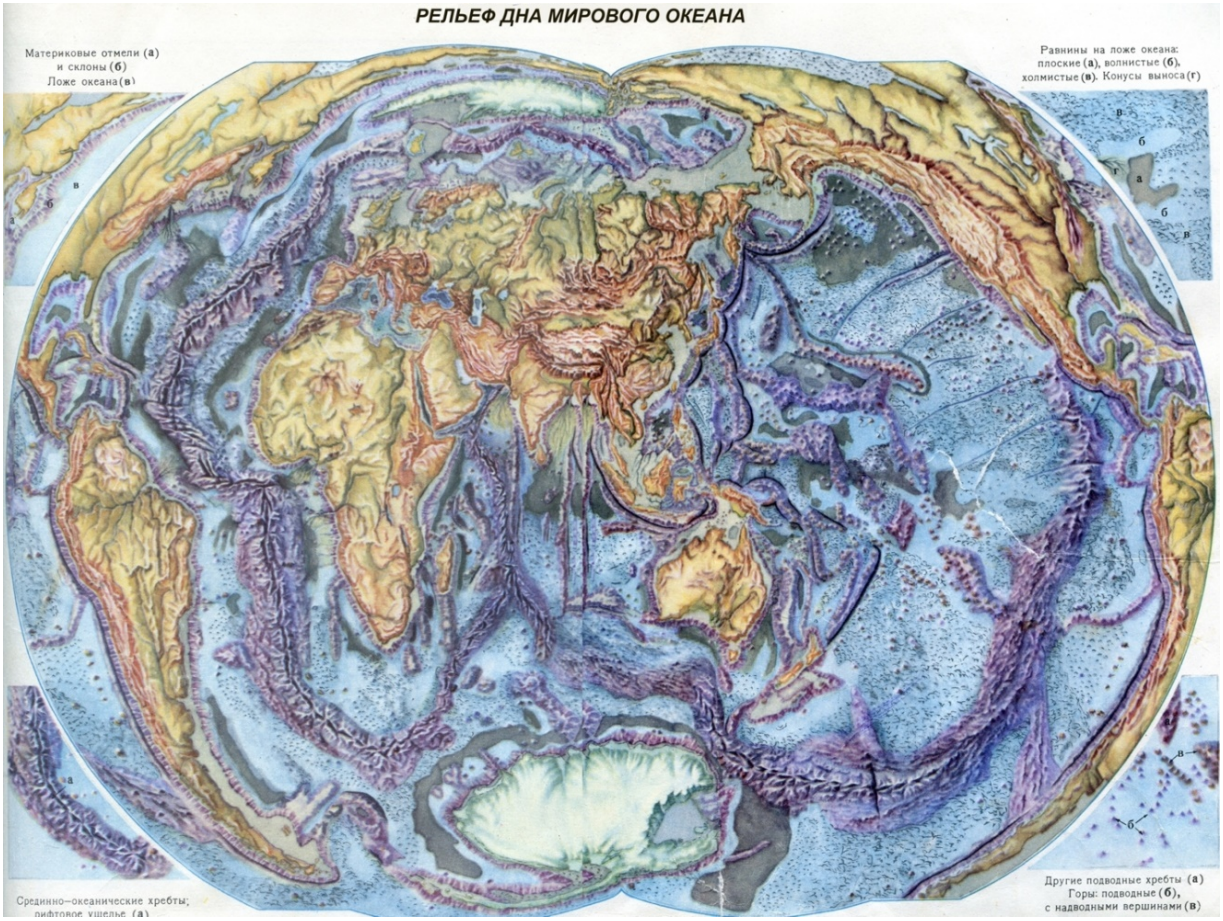


Рис. 1. Рельеф дна мирового океана

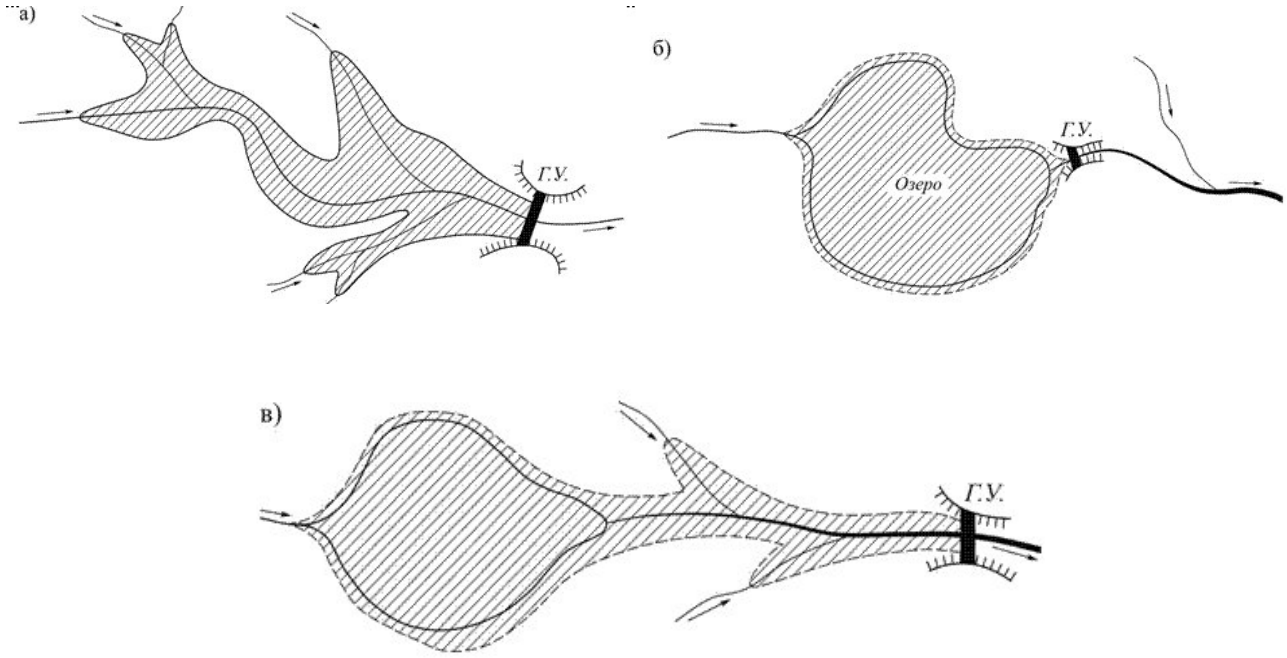


Рис. 2. Схемы слияния водных кластеров

## *Заключение*

Предлагаемая схема слияния кластеров, находит отражение в автоматизации алгоритмов управления гидрополивом заливных сельскохозяйственных полей, а также при отработке оперативного управления возведением защитных дамб при чрезвычайных событиях, вызванных наводнениями или грязевыми селями.

## *Благодарности*

Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0315-2016-0003).

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Ким П. А. Естественная кластеризация // Интерэкспо ГЕО-Сибирь. XIV Междунар. науч. конгр. : Междунар. науч. конф. «Дистанционные методы зондирования Земли и фотограмметрия, мониторинг окружающей среды, геоэкология» : сб. материалов в 2 т. (Новосибирск, 23–27 апреля 2018 г.). – Новосибирск : СГУГиТ, 2018. Т. 1. – С. 147–151.

2. Бучнев А. А., Ким П. А., Пяткин В. П., Пяткин Ф. В., Русин Е. В. Фреймворк сети облачных web-сервисов для программного комплекса обработки данных дистанционного зондирования // В сборнике: Региональные проблемы дистанционного зондирования Земли. Материалы V Международной научной конференции. Сибирский федеральный университет, Институт космических и информационных технологий. – 2018. – С. 7–11.

3. Ким П. А. О формализации модели естественной кластеризации // В сборнике: Региональные проблемы дистанционного зондирования Земли. Материалы V Международной научной конференции. Сибирский федеральный университет, Институт космических и информационных технологий. – 2018. – С. 128–130.

4. Водохранилища, их классификация и характеристики [Электронный ресурс]: https://helpiks.org/6-39388.html (дата обращения: 31.03.2019).